PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Inter-Hospital Comparison of Working Time Allocation Among Internal Medicine Residents Using Time-Motion Observations – An innovative benchmarking tool
AUTHORS	Frey, Simon Martin; Méan, Marie; Garnier, Antoine; Castioni, Julien; Wenger, Nathalie; Egloff, Michael; Marques-Vidal, Pedro; Beer, Juerg-Hans

VERSION 1 – REVIEW

REVIEWER	Lauren Block Northwell Health, USA
REVIEW RETURNED	26-Aug-2019

GENERAL COMMENTS	This is a time motion study of internal medicine residents completed at 2 hospitals in Switzerland. The goal of the paper is to determine whether time motion can be used to compare two residency programs. Since this is the goal, as the reader I expected the paper to be primarily methodological; for authors to detail the time motion procedure and its accuracy, including quality control of the observers and equivalency between weeks or shifts. Instead of focusing on the validity of the methodology and the pros and cons of the approach, the majority of the paper dwells on the specific differences between the hospitals. There seem to be multiple confounding factors which may have played a role in determining the differences between these hospitals, including resident experience, University status, use of doctor assistants, etc. I'm not sure given all these confounding variables what conclusions can be drawn from the fact that residents at Hospital B spend less time with patients than residents at Hospital A. I was hoping to glean whether time motion can be used to compare programs, with some confidence as to the validity of the methodology via use of quality assurance, but there was hardly any focus on this in the paper, which ultimately I find a major limitation given this is the primary research question. A few more minor questions below: p3: "Compared with self-declaration, the study design with peer-observers did less influence resident's allocation of time and reveals more objective (and precise) data" How can the authors be sure that use of peer observers less infuenced time allocation and revealed more objective data than a methodology that was not used in this study? Might qualify as 'might' or instead cite the benefits of a time motion approach

	p3 The authors cite that the Ethics Committees deemed the project to be exempt from review. It would be helpful to know which Ethics Committees reviewed this study and gave that recommendation. While the authors state that the study was exempt, later they specify that the residents gave written informed consent. Who reviewed and approved the consent forms if the study was exempt? p5 'For these tasks, they use computers and electronic medical records (EMR) up to 60%6,9,10' - do the authors mean up to 60% of the indirect time or the overall time? . p9 'Residents of hospital B, with a shorter average postgraduate training,' since the residents had less experience, this should be noted in the methods not explained in the discussion p9, 'However,the time spent for admission was more than twice as high at hospital A (28 vs. 11 minutes, p = 0.04) compared to hospital B.' is this the time spent actually doing the admission, or the time spent with each admitted patient? Were teams at both hospitals admitting an equivalent number of patients per shift? p10 ' In hospital A residents spent almost twice as much time together with colleagues when performing activities directly related to patients (134 vs. 76 min, p < 0.001).' what do the authors mean by 'time together with colleagues? does this mean in a resident room, or on rounds? Why was this outcome collected? The authors speculate in the introduction and the methods that delegating administrative tasks to other healthcare professionals might increase time available for direct patient care, yet in the study despite the fact that hospital B used doctor assistants, residents at hospital B spent less time in direct patient care. Since this hypothesis was not, in the end, shown to be the case in this study, I would ask the authors to comment on their thoughts as to the reasoning for this. Only day shifts were studied; constructing a limited picture of all patient care activities performed by residents. This may have been unavoidable given observer availability, but should be noted in the limitations.

REVIEWER	Lance Downing Stanford University, USA
REVIEW RETURNED	04-Sep-2019

GENERAL COMMENTS	I appreciate this significant effort by the authors to gain new insight into the work of medical trainees. I find this timely, important, and approached in a rigorous manner. However, some aspects of this manuscript could benefit from additional clarity. While the methods paper was referenced so perhaps out of scope, some elaboration on how activities were defined and the trade-offs between using other definitions would be helpful. Also, why were patients with private insurance excluded? This is not intuitive to this reviewer. It would also be helpful to clarify objective -- several are described and it is not clear what this study adds to the literature. It might also be helpful to review other methods of evaluating work activities such as EHR audit logs (published in several studies now).

VERSION 1 – AUTHOR RESPONSE

Answers to reviewer's 1 comments, Lauren Block:

1. **This is a time motion study of internal medicine residents completed at 2 hospitals in Switzerland. The goal of the paper is to determine whether time motion can be used to compare two residency programs. Since this is the goal, as the reader I expected the paper to be primarily methodological; for authors to detail the time motion procedure and its accuracy, including quality control of the observers and equivalency between weeks or shifts.**

   We would like to thank the reviewer for her in-depth revision of our manuscript.

   In order to follow the reader expectancy, we edited the manuscript and added more information regarding the methodology used. More specifically:

   1) We included more details on time motion procedures on page 6-9.
      a) We complemented the patient and public involvement section (details regarding changes are described under Reviewer Point 6)
      b) We extended the section "Data collection procedures" (page 6, line 296 – 334). The text now reads:
      "The study methods was described in details by Wenger et al[6]. Briefly, specially trained undergraduate medical students observed the resident's work during their shifts without interfering with the daily practice. The observer's training consisted of an e-learning program, in which they were introduced to the different definitions of activities. After this pre-training, they were trained in a hands-on session focusing on the definition of activities and the use of the recording device, followed by a 2-hour practice session based on a 1-hour training video, in which they observed a resident at work performing daily activities. For further training and quality control, we performed a blank recording of a resident's workday on a ward and a final session to resolve remaining issues. The inter-observer reproducibility to record residents' activities was assessed during these practice sessions within interactive discussions. After training, all observers recorded activities and contexts similarly. At hospital A, 6 observers were trained. Due to difficulties to cover all observation shifts at hospital B, 11 observers were trained."

      The position of the paragraph *"Additionally, we collected the baseline characteristics of residents including sex, age, country of medical school graduation, amount of postgraduate training and commuting distance to the hospital."* was changed from line 295 to 342.

   c) We created a new section "Recording device and definition of activities", inserted a new figure with the recording device (figure 2) and transferred the table with the definition of activities from the Appendix to the main document (former Appendix Table 1, current Table 2).

      The text was changed to the following: (page 7-9, lines 346 – 405)
      Activities were recorded with a tablet-based software designed and developed by the study team (figure 2). Details about the software can also be found in the methods section[6]. The software was designed to allow fast and easy changes between performed activities with a precision up to the second. To promote similar studies, the source code is available "as is" on the following github website (https: //github.com/agarnier00/MEDAY)[6].
      As the main goal for the development of this recording tool was to get the most precise data from the resident's workday, definitions covered the whole spectrum of tasks performed by residents. The observers were asked to record 22 different activities and their context as previously defined[6]. The same definitions were used at both hospitals (table 2).
      These 22 activities were divided in six categories (directly related to patients (n = 5), communication (n= 2, counted to directly related to patients as well), indirectly related to patients (n = 9), training (n = 3), non-medical tasks (n = 2) and transition (n = 1)). Definitions

of activities were specific to get a truthful picture of a resident's working day (e.g. activities directly related to patients differentiated admission, daily round, discharge, clinical procedures and out of unit support; indirect activities differentiated different tasks at the computer, supervision and more. See Table 2 for more details). All these activities could be recorded with one or more contexts present (with patient present, using phone, with colleague or using a computer). Because these contexts, especially time with patients, were of utmost importance, they could be recorded during all activities. To highlight the most important definitions: "patient present" was recorded when a resident was in direct contact with a patient and had his attention towards the patient. Reviewing laboratory results on the computer during daily round in the patient's room did not count for "patient present". Admission accounted for all admission activities (reading the record, anamnesis, physical examination, documentation), except during the daily round. Observers were teached to interpret local EMR system.
To be able to compare the amount of activities according to actual workload, the number of patients present at the observed ward was registered. To estimate the residents' workload on the day of observation, we calculated patient-equivalents on the basis of hours patients were present at the ward. One patient-equivalent was defined as the presence of a patient during the whole observed shift of an individual resident. Number of patients and hours of presence was extracted from the EMR."

2) Quality controls of the observers were performed using feedback techniques within interactive sessions. This was added on page 7, lines 329 – 335 (see 1a) for text changes). In the limitation section, we mentioned the lack of quantitative evaluation of inter-observer variability (page 15, lines 811 - 814).

   The text now reads:
   "Fourthly, we did not perform a quantitative analysis of inter-observer variability, but during extensive training sessions we checked correct judgement of activities on several occasions. After the training, observers reported similarly. "

3) We did not collect data during night shifts because the organization differed too much between hospitals and it was difficult to recruit motivated observers for night shifts. Therefore, we did not detail equivalency between weeks or shifts (mentioned in the limitation section on page 15, lines 806 - 811).

   "Thirdly, given the limited availability of observers during night shifts and the very different structural organization between hospitals during nights and weekends, only weekday shifts were considered for this study. As most clinical work is performed during dayshifts and only a small percentage of residents "guards" the hospital during evenings and nights, we estimate this lack of information as negligible."

4) To improve readability and give the reader the opportunity to better understand the different activities registered, we change the appearance of the figures and tables. We would therefore kindly ask the Editor to allow us to increase the actual number of tables and figures if possible:

a) We transferred the table with the definitions from the appendix to the main document (see 1c) for more details).
b) We inserted a screen shot (figure 2) of the tool used for data collection (see 1c) for more details).
c) We stated where to find the source code of this recording tool on page 7 (lines 350 – 362) in the methods section (see 1c) for text changes).
d) We transferred table 3 (Inter-hospital comparison of context in which direct or indirect activities were performed) to the Appendix. Is there a possibility to integrate this table also in the main document to enhance readability?

2. **Instead of focusing on the validity of the methodology and the pros and cons of the approach, the majority of the paper dwells on the specific differences between the hospitals.**

We added and edited a section entitled **"Time motion-studies to benchmark resident's allocation of time: Pros and cons"** in the discussion part (page 11-13, lines 516 - 630). We thank the reviewer for the idea.
Overall, we shortened the abstract, results, and discussion sections regarding specific differences between hospitals in order to focus more on the methodology.

The new section reads:
Time motion-studies to benchmark resident's allocation of time: Pro and cons

"In this study aiming to compare the allocation of time of residents at work in different hospital settings, we first showed that time-motion observations allowed to compare complex day structures of internal medicine resident's working in different hospital settings, i.e. a French-speaking university hospital and a German-speaking non-university hospital, with plausible and comparable results. Sinsky and associates[9] described the allocation of physicians' time in ambulatory practice. Similarly to us, they describe allocation of physician's working time in different settings and specialties (i.e. family medicine, internal medicine, cardiology and orthopedics) and concluded that factors present in one setting (for e.g. dictation services and documentation assistant services) might have influenced the direct clinical face time.
Second, time-motion studies are not time-consuming for the workers, i.e. the residents. As residents usually operate at full capacity, they have no time left for documenting their tasks performed precisely as this disrupts their workflow. Moreover, documenting the activities performed after the shift may result in a substantial loss of information (recall bias). Therefore, data collection using observers is one of the major advantages of time-motion studies, especially when the device is fast and flexible as the one used in this study (figure 2).
Thirdly, as time-motion studies give excellent information about the type and distribution of work activities, it can also be a well-suited method to compare post interventional changes in a single institution (such as change of EMR, increase of the number of staff, implementation of doctor's assistants). Additionally, this methodology could easily be adapted to examine the day structures of other health care professionals. However, because multiple factors can play a role in residents' time allocation and in fine determine the differences observed between hospitals, time-motion studies must be thought as a first step on the path for future changes. Indeed, based on time-motion results, each hospital will be responsible to test several interventions targeting residents' time allocation in order to modify resident allocation of time. Focus groups could help to find explanations regarding differences in residents' allocation of time.
Fourthly, as many tasks are complex and can possibly occur as "multitasks", time motion studies using devices allowing to record several tasks at the same moment might be of interest, as previously published[7]. However, in our group, we studied task-switching, defined as a switch from one task to another, expressed as number of switches per hour rather than multitasking[15] as from literature there is evidence, that the human brain can perform one or maximally two demanding tasks at the time, which are thought to be performed sequentially[16].
Fifthly, with the continous tracking of activities with a time stamp, time-motion studies can be used to generate heatmaps. This graphical presentation of data can help to detect patterns of time allocation and display the distribution of activities over the working day. For readers, this visual help can be more illustrative than percentage quotation in a table.

Sixthly, compared to self-reported surveys, in which residents could (in)voluntarily overestimate activities (i. e. patient visit, as residents might consider this task as the most important), time-motion studies objectively describe time allocation and point out potential differences between settings. However, self-report survey remain less time-consuming for reseachers or administrators and cheaper.
Indeed, financial ressources required for time-motion studies are costly, but affordable for most hospitals or divisions and therefore usable. For example, an observation period of 500 hours at a hourly rate salary of 20 Swiss Francs (CHF) results in 10'000 CHF for the observation period only. However, compared to self-declaration studies, their logisitical and financial costs are much higher. Another interesting alternative to time motion-studies can be Electronic Health Records (EHR) audit logs. In these analysis of data protocols from EHRs a huge amount of data can be extracted without intervening with the daily routine of residents (or other physicians). Patterns in the working day can be detected and workflow of different professionals can be compared. But compared to time-motion studies, this method lacks important information on the full spectrum of clinical activities which are not performed with the mouse or the keyboard.
In conclusion, time-motion studies are interesting and innovative hospital benchmarking tools able to demonstrate how physicians allocate their working time in different settings or culture."

We clarified also the objectives (page 4-5, lines 190-252) :
Thus, the first study goal was to describe how this method could be used for inter-hospital comparison. The secondary goal was to learn more about specific differences in the resident's working day structure in university and non-universtiy hospital settings. Because studied hospitals have differing teaching objectives and organization, we hypothetized that we would find differences in residents's allocation of time.

3. **There seem to be multiple confounding factors which may have played a role in determining the differences between these hospitals, including resident experience, University status, use of doctor assistants, etc. I'm not sure given all these confounding variables what conclusions can be drawn from the fact that residents at Hospital B spend less time with patients than residents at Hospital A.**

We agree with the reviewer that multiple factors can play a role in residents' time allocation and, *in fine,* could determine the differences observed between hospitals. Actually, we stated it in the abstract and it was our working hypothesis. Thus, to make it clearer, we added in the introduction section (page 5, lines 250-252):
"Because studied hospitals have differing teaching objectives and organization, we hypothetized that we would find differences in residents' allocation of time."

Moreover, in a time-motion study, published in the Annals of Medicine, Sinsky and associates[1] described the allocation of physicians' time in ambulatory practice. Similarly to us, they compared allocation of physician's working time in different settings and specialties (i.e. family medicine, internal medicine, cardiology and orthopedics). They concluded that factors present in one setting (for e.g. speech recognition) might have influenced the time dedicated to administrative tasks. We discuss this point within the section pro and cons on page 11 (lines 517-526) and cited the above mentioned citation (see Reviewer Point 2).

Thus, time-motion studies are able to objectively point out differences in time allocation, however without proof of causality, that will pave the path for future changes. Based on time-motion results, each hospital will be responsible to test several interventions targeting residents' time allocation in order to modify resident allocation of time. We are currently setting up focus groups to find explanations regarding difference in residents' allocation of

time.

This study limitation was already stated it the limitation section (page 14, lines 803 - 806):
"Second, because of the observational study design, it could not be assessed whether certain organizational factors, such as doctor's assistants, various EMR systems, different language or culture had a causal effect on time allocation."

We added the point about needs of intervention or qualitative studies on page 16 (lines 848 - 852):
"Additionally, using doctor's assistants could lead to an increase in workflow efficiency as shown by Bank and colleagues[2]. Further studies in the same hospital are needed. Finally, qualitative analyses, such as focus groups, could determine the reasons why residents allocate their time in such manners and add informative data."

4. **I was hoping to glean whether time motion can be used to compare programs, with some confidence as to the validity of the methodology via use of quality assurance, but there was hardly any focus on this in the paper, which ultimately I find a major limitation given this is the primary research question.**

Thanks for the comment, we tried as specified above to focus on the methodology.

5. **p3: "Compared with self-declaration, the study design with peer-observers did less influence resident's allocation of time and reveals more objective (and precise) data" How can the authors be sure that use of peer observers less influenced time allocation and revealed more objective data than a methodology that was not used in this study? Might qualify as 'might' or instead cite the benefits of a time motion approach**

We agree with the reviewer that this statement is beyond the goal of our study. Therefore we changed the text on page 3 (lines 123-125) in:

"A time-motion study design with peer-observers produces objective and precise data as residents do not use their subjective resources for data acquisition and are only minimally disturbed by the observers."

6. **p3 The authors cite that the Ethics Committees deemed the project to be exempt from review. It would be helpful to know which Ethics Committees reviewed this study and gave that recommendation.**

We specified that the Human Research Ethics Committees of both counties (Nordwestern and Central Switzerland and Canton Vaud) certified that there is no need for ethical approval (documents available on request), as the focus of the study was on the residents and not on patients with the aim to improve the quality of resident work, and by extension the quality of care. The text now reads (page 6, lines 288 - 294):

"All residents gave written informed consent to use their anonymized data for quality improvement purposes and signed an informed consent form. The local study team verified that written informed consent was present from all participants before planning the

observation. As no patient data or health information and no data on residents' health were used for this study, the Human Research Ethics Committees (Northwestern and Central Switzerland and Canton Vaud) certified that the study was exempt from human subjects ethics review and therefore needed no approval."

7. **While the authors state that the study was exempt, later they specify that the residents gave written informed consent. Who reviewed and approved the consent forms if the study was exempt?**

Indeed, residents were followed during their normal daily work as employees. Any quality control (or supervision) of processes in a company could be decreed from the employer. To insure residents that the data will be used for quality of work improvement purposes only and not for human resource purposes and that results will have no consequences for them, we only included residents who gave written informed consent. The local study team controlled that informed consent was present before planning the observation of the corresponding resident.

We rephrased it in the section "Patient and public involvement" on page 6 (see Reviewer point 6).

8. **p5 'For these tasks, they use computers and electronic medical records (EMR) up to 60% [6,9,10]' - do the authors mean up to 60% of the indirect time or the overall time?**

Thanks for the remark. We specified it:

"For these indirect tasks, they use computers and electronic medical records (EMR) up to 60% of the time for indirect task[6,9,10]." (page 4, 170-171)

9. **p9 'Residents of hospital B, with a shorter average postgraduate training,' since the residents had less experience, this should be noted in the methods not explained in the discussion**

This information was cited in Table 1 and mentioned in the method section on page 5. However, we deleted the information "on average PG training" in the result section. The text now reads:

"Residents of hospital B required more time to look up patient specific information (51 vs. 38 minutes, p = 0.022)." (page 10, 441 - 442)

10. **P8, 'However, the time spent for admission was more than twice as high at hospital A (28 vs. 11 minutes, p = 0.04) compared to hospital B.' is this the time spent actually doing the admission, or the time spent with each admitted patient?**

According to the study definitions, 28 or 11 minutes are the average time recorded as "doing the admission". To help the reader understand these definitions, we moved the table

containing study definitions of activities from Appendix Table 1 to the main article Table 2 (page 8).

However, we agree with the reviewer, the amount of time dedicated to admission is low. Time dedicated to admission was probably reported in the activity entitled “daily rounds” since some newly admitted patients are “discovered” by residents in the morning during the round. Because observers must stick to the 22 definitions, some activities occurring in an unexpected setting or time frame may be recorded as another activity, within the same category i.e. “activities directly related to the patient”.

We added the risk of misclassification as a limitation (page 15, 819 - 823):

“Seventhly, some activities (for e.g. patient’s admission) occurring in an unexpected setting or time frame might have been recorded as another activity (for e.g. daily round). However because it remains within the same category (i.e.“activities directly related to the patient”) we doubt such misclassification could significantly bias the results.”

We also changed the wording on page 10 (lines 442 – 444) to clarify that the values are for each resident and not per patient:

“However, the time each resident spent for admissions was more than twice as high at hospital A (28 vs. 11 minutes, p = 0.04) compared to hospital B.”

11. **Were teams at both hospitals admitting an equivalent number of patients per shift?**

The number of admissions per resident and per shift observed was not collected. To compare residents’ workload, we calculated patient-equivalents which is defined as follows: One patient-equivalent was defined as the presence of a patient during the whole observed shift of an individual resident. Number of patients and hours of presence was extracted from the EMR.

We adapted the text on page 8 (lines 385 – 405) accordingly:
“To be able to compare the amount of activities according to actual workload, the number of patients present at the correponding ward was registered. To estimate the actual residents’ workload on the day of observation, we calculated patient-equivalents on the basis of hours patients were present at the ward. One patient-equivalent was defined as the presence of a patient during the whole observed shift of an individual resident. Number of patients and hours of presence was extracted from the EMR.”

We also added the following limitation on page 15 (lines 823 - 825):
“Fifthly, no data on the number of admissions and discharges per shift was collected, but we estimated the workload based on the number of patients each resident was in charge.”

12. **p9 ' In hospital A residents spent almost twice as much time together with colleagues when performing activities directly related to patients (134 vs. 76 min, p < 0.001).' what do the authors mean by 'time together with colleagues? does this mean in a resident room, or on rounds? Why was this outcome collected?**

As stated in the Table 2 (former Appendix eTable 1), we recorded 22 different activities. All these activities can be recorded with one or multiple "context" items. This allows to reveal more insights about the activities performed: for e.g. when a resident interacts with a patient, the context "with a colleague" permits to know whether he did alone or with someone and for how long.

We added this information in the method section on page 8 (375 – 378): *"All these activities could be recorded with one or more contexts present (with patient present, using phone, with colleague or using a computer). Because these contexts, especially time with patients, were of utmost importance, they could be recorded during all activities."*

We also rephrased the text in the results section on page 10 (451 – 453):

"When residents in hospital A performed activities directly related to patients, there was more often a colleague present (134 vs. 76 min, p < 0.001)."

13. **The authors speculate in the introduction and the methods that delegating administrative tasks to other healthcare professionals might increase time available for direct patient care, yet in the study despite the fact that hospital B used doctor assistants, residents at hospital B spent less time in direct patient care. Since this hypothesis was not, in the end, shown to be the case in this study, I would ask the authors to comment on their thoughts as to the reasoning for this.**

Indeed Bank and colleagues (Bank et al, CEOR, 2013[2]) showed in an ambulatory setting that medical scribes (another form of doctor's assistants) increased the average time physicians spent interacting with patients from 1.5 to 6.7 minutes per visit while the visit time was shorter (9.1 vs. 12.9 minutes). Thus, we agree with the reviewer that delegating administrative tasks ended in lower time dedicated to administrative task without increasing the average time dedicated to patients. As comparing two sites, we can not proof causality.

We therefore changed the text accordingly (page 14, lines 646 - 650):
"Interestingly, residents in the non-university hospital spent less time performing administrative activities (18 vs. 32 min, p < 0.001) and answering the phone (41 vs. 58 min, p = 0.021), which might be an effect attributable to the available doctor's assistants. This is partially in accordance with our study hypothesis and the literature[2]."

We changed our statement in the implication section:
"On the basis of our results, several interventions targeting residents' time allocation could be tested in hospital practice. First, delegation of administrative or low added value tasks to doctor's assistants (as it was done in hopistal B) could allow residents to focus on more valuable medical activities and training (as shown in the ambulatory setting by Bank et al.[12]). Second, optimizing documentation supports, using speech or writing recognition systems, or improving EMR ergonomy could decrease the time residents spend writing in the EMR[9]. Additionally, using doctor's assistants could lead to an increase in workflow efficiency as shown by Bank and colleagues[12]. Further studies in the same hospital are needed. Finally, qualitative analyses, such as focus groups, could determine the reasons why residents allocate their time in such manners and add informative data." (page 15-16, lines

842 - 852)"

14. **Only day shifts were studied; constructing a limited picture of all patient care activities performed by residents. This may have been unavoidable given observer availability, but should be noted in the limitations.**
We thank the reviewer for this legitimate remark and adapted the text accordingly in the limitation section (page 15, 806 - 811).

"Thirdly, given the limited availability of observers for night shifts and very different structural organization between hospitals during nights and weekends, only weekday shifts were considered for this study. As most clinical work is performed during dayshifts and only a small percentage of residents "guards" the hospital during evenings and nights, we estimate this lack of information as negligible."

15. **I appreciate this significant effort by the authors to gain new insight into the work of medical trainees. I find this timely, important, and approached in a rigorous manner.**

We thank the reviewer for this comment.

16. **However, some aspects of this manuscript could benefit from additional clarity. While the methods paper was referenced so perhaps out of scope, some elaboration on how activities were defined and the trade-offs between using other definitions would be helpful.**

We thank the reviewer for this remark and adapted the methods part accordingly and transferred the Appendix eTable 1 with all definitions used to the main document (Table 2). See also Reviewer Point 1 for more details.

17. **Also, why were patients with private insurance excluded? This is not intuitive to this reviewer.**

In Switzerland, private insured patients' wards have usually a different workflow for the resident compared to normal wards. Moreover, private patient in Switzerland benefit from single room and have the choice of their attending physician, which potentially is not the case in other countries. Therefore, we preferred to focus on comparable patients wards to enhance the generalizability of our results.

We added it in the limitation section on page 15 (lines 823 - 825): *"Finally, in order to enhance the generalizability of our results, we did not observe residents working in privately insured patient wards, because private insurance systems are country-dependent."*

18. **It would also be helpful to clarify objective -- several are described and it is not clear what this study adds to the literature.**

We rephrased the objectives in the last paragraph of the introduction (page 4, 182 – 252):

"Recently Sinsky et al [9] performed a time-motion study including different specialities (medical and surgical) in different settings. This work gave first insights of physicians' time allocation in various ambulatory settings. In the hospital practice, many factors such as hospital type (university vs. non-university), patients length of stay and amount of comorbidities (case-mix index), local educational culture and habits may influence the resident's working day structure, we therefore aimed to compare working time allocation among internal medicine residents using time-motion observations, in two large Swiss hospitals, using a method developed previously[6].

Thus, the first study goal was to describe how this method could be used for inter-hospital comparison. The secondary goal was to learn more about specific differences in the resident's working day structure in university and non-universtiy hospital settings. Because studied hospitals have differing teaching objectives and organization, we hypothetized that we would find differences in residents's allocation of time."

We also edited the discussion section in order to discuss one objective after the other.

19. **It might also be helpful to review other methods of evaluating work activities such as EHR audit logs (published in several studies now).**
We thank the reviewer for the relevant suggestion and added a paragraph in the discussion part (page 13, 620 - 627). See also Reviewer Point 2 for the text changed.

Reference

1. Sinsky C, Colligan L, Li L, et al. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. *Ann Intern Med*. 2016;165(11):753-760. doi:10.7326/M16-0961

2. Bank AJ, Obetz C, Konrardy A, et al. Impact of scribes on patient interaction, productivity, and revenue in a cardiology clinic: a prospective study. *Clinicoecon Outcomes Res*. 2013;5:399-406. doi:10.2147/CEOR.S49010

VERSION 2 – REVIEW

REVIEWER	Lauren Block Zucker School of Medicine, USA
REVIEW RETURNED	05-Nov-2019

GENERAL COMMENTS	Thank you for your much improved revised version of this time motion study of internal medicine resident activities on general medicine ward services at two hospitals in Switzerland. I found this version to be much more readable and clear. The implications section as well as the heatmaps were extremely helpful in highlighting potential lessons learned to optimize time spent with patients. A few questions remain: 1. On p7 "After training, all observers recorded activities and contexts similarly" was there an interrater reliability score

	calculated? This would be helpful to describe the similarity between observers. 2. Also p7, separate observers were used in the two hospitals. Were these observers trained together? Was their accuracy similar in training sessions? It would be helpful to know if any differences seen was due to differences between the observers. 3. p8 "Observers were teached" would switch wording to 'taught. 4. Also p8 could the observers record more than one activity simultaneously? 5. p9 what % of residents consented to participate at each site? 6. p14 why do the authors posit that a less experienced group was recruited at Hospital B? This seems to be a significant potential confounder since the authors attribute differences in efficiency to differential experience. 7. Also p14 it is surprising that despite the presence of doctors assistants at Hospital B to help with administrative tasks, the residents spent considerably less time with patients. The authors seem to speculate that less experience and therefore less efficiency was the reason; are there other possible explanations? Is the day structured differently (for example, different length of teaching or rounds) which may be partially responsible for the differences seen? 8. The authors state that the current methodology can be used to compare residents across hospitals. Would hospitals with less time spent with patients (ie Hospital B) have concern about these comparisons? 9. p16 Why were only daytime shifts observed? This might be added to the Limitations section. 10 Small point but I found it surprising that computers were not used during handoffs.

REVIEWER	Lance Downing Stanford University, USA
REVIEW RETURNED	06-Nov-2019

GENERAL COMMENTS	Interesting application of time and motion methodology that has previously been used to examine physician work activities. Interesting innovation using peer reviewers and tablet based platform to understand trainee work activities and differences between hospitals. This seems like a very promising approach and expect this to be a platform for more nuanced studies of how clinicians work.

VERSION 2 – AUTHOR RESPONSE

Answers to reviewer's 1 comments, Lauren Block:

Thank you for your much improved revised version of this time motion study of internal medicine resident activities on general medicine ward services at two hospitals in Switzerland. I found this version to be much more readable and clear. The implications section as well as the heatmaps were extremely helpful in highlighting potential lessons learned to optimize time spent with patients. A few questions remain:

1. On p7 "After training, all observers recorded activities and contexts similarly". Was there an interrater reliability score calculated? This would be helpful to describe the similarity between observers.

Thanks for giving us the opportunity to give more details on the methods.

Due to different study time points and language (French /German), observers differed between centers and were trained separately. At hospital A, 6 French-speaking observers were trained. At hospital B, 11 German-speaking observers were trained. The training was similar in both hospital and is described below.

In hospital A the reproducibility of all observers was assessed during the video practice. (See table, which is also published in the Appendix of this publication[1]). In hospital B, the reproducibility of a sub-group of observers was assessed (n=5).

We now mentioned the range of coefficient of variation for hospital A and B and stated the lack of complete quantitative assessment in hospital B (n=5 out of 11 observers) in the limitations section (page 15, line 25).

“Fourthly, the observer reproducibility was assessed for a sub-group of observers in Hospital B (5 in 11) versus for all 6 observers in Hospital A. This method is partially in accordance with a previous time-motion study[2]. Moreover, coefficients of variation were smaller than 10% for categories of activities in both hospitals.”

Table: results of the reproducibility studied between observers of Hospital A and B during the video practice

Hospital A	**Minimum**	**Maximum**	**Average**	**CV (%)**
Overall	68.7	69.3	68.9	0.3
Categories				
Directly related	12.0	12.2	12.1	0.7
Indirectly related	36.8	37.9	37.5	1.3
Academic	1.2	1.5	1.4	9.1
Non-medical	17.4	18.9	17.9	3.3
Contexts				

Patient	0.5	0.9	0.6	29.9*
Computer	46.1	50.8	48.4	3.6
Telephone	3.9	6.1	5.1	17.1
Colleague	14.3	19.1	15.8	12.4

Hospital B	**Minimum**	**Maximum**	**Average**	**CV (%)**
Overall	42.7	43.5	43.0	0.6
Categories				
Directly related	14.9	17.1	16.0	4.9
Indirectly related	13.2	14.3	13.7	3.4
Academic	3.0	3.9	3.6	9.3
Non-medical	9.2	10.9	9.6	7.4
Contexts				
Patient	4.4	5.9	5.5	11.4
Computer	16.5	20.9	18.6	8.3
Telephone	1.1	2.8	1.6	46.9*
Colleague	11.7	13.4	12.6	5.0

Results are expressed in minutes. CV: coefficient of variation. * The values are high due short periods recorded. Number of observations (hospital A: n = 6, hospital B: n = 5).

The text reads now as follows: (page 6, line 22)

"Briefly, specially trained undergraduate medical students observed the resident's work during their shifts without interfering with the daily practice. Due to different study time points and local language, observers were trained at each center separately. At hospital A, 6 French-speaking observers were trained. At hospital B, 11 German-speaking observers were trained. The training was similar in both hospitals and consisted in a) a dedicated e-learning program on how to categorize the various resident's activities; b) a teaching session, focused on the definition of activities and the use of recording device; c) a two hours practice session based on a video of residents engaging in typical medical activities; d) eight hours of blank observation to observe and record a resident on the ward, and e) a

last session to control they correctly recorded activities during blank observation, to point out specific situation and to standardize the observations. In hospital A the reproducibility of the 6 observers was assessed during the video practice for all observers. Overall, observers recorded activities similarly. The coefficient of variation for categories ranged between minimum 0.7 % for activities directly related to patients to maximum 9.1% for academic activities[6]. In hospital B, the reproducibility of a sub-group of observers (n=5) was assessed. The coefficient of variation ranged from minimum 3.9% for activities indirectly related to patients to maximum 9.3% for academic activities. Overall, coefficients of variation for categories of activities were quite similar and smaller than 10%, meaning that the observers in hospital A and B had similar accuracy during training sessions to record residents' activities."

2. Also p7, separate observers were used in the two hospitals. Were these observers trained together?

See our answer above.

3. Was their accuracy similar in training sessions? It would be helpful to know if any differences seen was due to differences between the observers.

Based on the results above, accuracy was good and similar in both hospitals.

We specified it in the method section (page 7, line 14-17):

"Overall, coefficients of variation for categories of activities were quite similar and smaller than 10%, meaning that the observers in hospital A and B had similar accuracy during training session to record residents' activities."

3. p8 "Observers were teached" would switch wording to 'taught.

We thank the reviewer for this correction and adapted it.

4. Also p8 could the observers record more than one activity simultaneously?

No. The software allows only one of these 22 activities to be recorded at the time, but different contexts (with patient, computer, phone, colleague) could be recorded concomitantly and could change irrespective to the activity being performed.

We specified with more details, page 8, line 21:

“All these activities could be recorded concomittantly with one or more contexts present (with patient present, using phone, with colleague or using a computer). Because these contexts, especially time with patients, were of utmost importance, they could be recorded during all activities. The tool allowed only one activity to be recorded at the time whereas the different contexts (with patient, computer, phone, colleague) could change irrespective to the activity being performed.”

5. p9 what % of residents consented to participate at each site?

At hospital A, 100% (n = 28) of residents, who were working at the observed wards at the study time point, consented for the study. At hospital B, 95 % (21/22) of residents, who were eligible, consented for the study.

We adapted the text accordingly page 10, line 4:

“In hospital A, all residents eligible for the study consented (100%, 28/28), whereas one resident in hospital B refused to participate (95% 21/22).”

6. p14 why do the authors posit that a less experienced group was recruited at Hospital B? This seems to be a significant potential confounder since the authors attribute differences in efficiency to differential experience.

In Switzerland, residency programs most frequently start in non-university hospitals. Residents in third post-graduated years start to work in university hospitals. Therefore, this difference in clinical experience reflects the “real world” situation. Pointing out difference in time allocation, even due to resident experience, could be of interest for medical educator, hospital administrators, or even for patients if they can choose their hospital.

We adapted the text in the limitations section, page 15, line 14:

“Second, because of the observational study design, it could not be assessed whether certain organizational factors, such as doctor’s assistants, various EMR systems, different language or culture had a causal effect on time allocation. Moreover, because residency in Switzerland starts in non-university hospitals, residents were less experienced at hospital B, which could have influenced their time allocation.”

7. Also p14 it is surprising that despite the presence of doctor’s assistants at Hospital B to help with administrative tasks, the residents spent considerably less time with patients. The

authors seem to speculate that less experience and therefore less efficiency was the reason; are there other possible explanations? Is the day structured differently (for example, different length of teaching or rounds) which may be partially responsible for the differences seen?

We thank the reviewer for this critical comment.

We also think that the difference in daily round (180 min vs. 100 min) is indeed probably the most relevant reason for the difference in time spent with patients. Residents at hospital A spent 139 minutes (from the planned 180 minutes) on daily rounds whereas residents in hospital B 86 minutes (from 80 minutes planned). As overall working time is an issue for hospitals as the maximum weekly hours are limited by law, daily round was limited a few years before at hospital B and doctor's assistants were implemented to reduce excessive working hours in residents. This led to shorter overall working time but can explain why residents spent less time with patients.

We adapted the text on page 14, line 20:

"Although residents from the non-university hospital spent less time performing administrative tasks, they dit not convert this advantage in time allocated to patients. As illustrated in the heatmaps (figure 3, panel D), most patient contact occured during daily round. Thus, because daily round was planned to last 180 minutes in the university hospital versus 80 minutes in non-university hospital (see figure 1), the potential beneficial effect of doctor's assistants could not be converted in more time with patients. "

8. The authors state that the current methodology can be used to compare residents across hospitals. Would hospitals with less time spent with patients (ie Hospital B) have concern about these comparisons?

This is a good question.

Whether more time with patients during an acute hospitalization means a better quality of care remains unsure, thus we do not think that our results should raise concerns.

Less time dedicated to patients decreases physician's satisfaction[3], patient education and health promotion[4], and increases inappropriate prescribing and medical malpractice.[5] However, most of this literature comes from the ambulatory setting.

Quality of care in hospital setting is probably related to how much time the patient spends with doctors (senior staff physician, specialist, medical student, specialized nurses, …) per day during an acute hospital stay. Importantly, our study reflects the viewpoint of the resident.

Moreover, one should not focus only on absolute minutes the resident spent per patient, but also on what was done in these minutes. If for example doctor's assistants or patient managers take over tasks which can be delegated (organizing discharge with e.g. ambulatory nursing care, explaining medication plan, performing simple examinations (e.g. ankle-brachial index)), this time can be used for other high value tasks such as discussion of the case with a senior, studying the history thoroughly and reviewing the literature for the correct treatment. Such tasks do not necessarily have to be performed at the bedside but could impact the quality of care.

We think that this discussion is beyond the goal of our study, therefore we did not add it all to the text.

We added as last sentence, page 16, line 26:

"Also, further studies additionally considering outcomes of patients, quality of care and patient / physician satisfaction would be helpful to better understand and optimize hospital workflows."

9. p16 Why were only daytime shifts observed? This might be added to the Limitations section.

We agree. It was written on page 15, line 19:

"Thirdly, given the limited availability of observers during night shifts and the very different structural organization between hospitals during nights and weekends, only weekday dayshifts were considered for this study. As most clinical work is performed during weekday dayshifts and only a small percentage of residents "guards" the hospital during evenings and nights, we estimate this lack of information as negligible."

10. Small point but I found it surprising that computers were not used during handoffs.

This is likely due to 1) hospital setting, which does not facilitate the usage of the computers (no tablets, only desktops) while interacting with colleagues and 2) the reluctance of residents to use the computer during handoffs, handoffs which are more and more performed using a TEAM-STEPPS[6] structure that computer use would slow down.

Answers to reviewer's 2 comments, Lance Downing:

Interesting application of time and motion methodology that has previously been used to examine physician work activities. Interesting innovation using peer reviewers and tablet

based platform to understand trainee work activities and differences between hospitals. This seems like a very promising approach and expect this to be a platform for more nuanced studies of how clinicians work.

We thank the reviewer for this comment.

References:

1. Wenger N, Mean M, Castioni J, Marques-Vidal P, Waeber G, Garnier A. Allocation of Internal Medicine Resident Time in a Swiss Hospital: A Time and Motion Study of Day and Evening Shifts. *Ann Intern Med*. 2017;166(8):579-586. doi:10.7326/M16-2238

2. Chaiyachati KH, Shea JA, Asch DA, et al. Assessment of Inpatient Time Allocation among First-Year Internal Medicine Residents Using Time-Motion Observations. *JAMA Intern Med*. 2019;179(6):760-767. doi:10.1001/jamainternmed.2019.0095

3. Shanafelt TD, Dyrbye LN, Sinsky C, et al. Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional Satisfaction. *Mayo Clin Proc*. 2016;91(7):836-848. doi:10.1016/j.mayocp.2016.05.007

4. Wilson A, McDonald P, Hayes L, Cooney J. Health promotion in the general practice consultation: a minute makes a difference. *BMJ*. 1992;304(6821):227-230.

5. Dugdale DC, Epstein R, Pantilat SZ. Time and the patient-physician relationship. *J Gen Intern Med*. 1999;14 Suppl 1:S34-40.

6. Starmer AJ, O'Toole JK, Rosenbluth G, et al. Development, Implementation, and Dissemination of the I-PASS Handoff Curriculum: A Multisite Educational Intervention to Improve Patient Handoffs. *Acad Med*. 2014;89(6).

VERSION 3 – REVIEW

REVIEWER	Lauren Block Northwell Health
REVIEW RETURNED	07-Dec-2019

GENERAL COMMENTS	Thank you for the opportunity to review this revised manuscript. I appreciate the authors' attention to detail and edits to address several of the concerns raised regarding an earlier version of the paper, including a description of the similar training program at both hospitals and an assessment of the concordance of observations. 1. The authors included a description of the concordance of observations among the observers at Hospital A and a sample of 5

	observers at Hospital B. How was the sample of observers at Hospital B chosen? Concordances between observers at Hospital A and Hospital B are reported; were these concordances compared to each other or to a 'gold standard'? This would help ensure consistency of observation between the two sites, which is important as this is the main comparison of the paper. 2. Thank you for the inclusion on page 14 of the fact that although residents at the non-University hospital spent less time on administrative tasks, this was not converted into more time with patients. It would be helpful to learn the authors reasoning for this, including varying experience level as noted on page 15. 3. Final sentence before conclusion on page 17 is awkwardly written; would consider rephrasing. Thank you for the change to review this much improved revision and look forward to seeing this in print.

VERSION 3 – AUTHOR RESPONSE

Answers to reviewer's 1 comments, Lauren Block:

Thank you for the opportunity to review this revised manuscript. I appreciate the authors' attention to detail and edits to address several of the concerns raised regarding an earlier version of the paper, including a description of the similar training program at both hospitals and an assessment of the concordance of observations.

1. The authors included a description of the concordance of observations among the observers at Hospital A and a sample of 5 observers at Hospital B. How was the sample of observers at Hospital B chosen?

Pragmatically, only 5 recording devices were available at hospital B, thus it determined the final number of observers at the time. Within the 11 observers at hospital B, the sample of 5 was chosen at random.

We added the following statement in the text (limitation section, p16, line 1-4):

"Fourthly, not all observers could be assessed for observer reproducibility as at hospital B, only 5 recording devices were available, which determined the final number of observers assessed at random for reproducibility."

Concordances between observers at Hospital A and Hospital B are reported (in the second rebuttal letter); were these concordances compared to each other or to a 'gold standard'? This would help ensure consistency of observation between the two sites, which is important as this is the main comparison of the paper.

We are not aware of a "gold standard" to assess activity time and we would be very glad if the reviewer could provide us some references of similar studies that used such gold standard. The videos used in both hospitals might be considered as a proxy for the gold standard.

We added the following statement in the discussion, limitations (p16, line 15:

"Further, no gold standard (for the correct activity) was available regarding the assessment of the different activities; hence, a systematic reporting bias (i.e. the observers from one hospital systematically over or underestimating a given activity) cannot be completely ruled out. Still, this issue occurred in other studies as well [7,9] and it would be important that a common gold standard be created to facilitate comparison between studies and settings."

2. Thank you for the inclusion on page 14 of the fact that although residents at the non-University hospital spent less time on administrative tasks, this was not converted into more time with patients. It would be helpful to learn the authors reasoning for this, including varying experience level as noted on page 15.

We added the following statement on page 14:

"Moreover, because residency in Switzerland starts in non-university hospitals, residents were less experienced at hospital B, which could have influenced their time allocation, in particular the time dedicated to patients. One can also argue that less experienced residents may favour "iPatient" care, i.e. using the EHR as the virtual construct of the patient, in order to avoid asking time-consuming questions."

3. Final sentence before conclusion on page 17 is awkwardly written; would consider rephrasing.

Thanks. We rephrased (p 17, lines 11-13):

"To better understand and optimize hospital workflows, it would be helpful to integrate patient outcomes, quality of care related-outcomes and satisfaction outcomes in future time-motion studies."